oxide synthases (NOSs) is achieved by catalysis of the two-step oxidation of L-arginine to L-citrulline. As in ▶ cytochrome P450 enzymes, the sulfur of a cysteine residue coordinates to the heme iron. In addition, NOSs feature a conserved tryptophan that forms a hydrogen bond with the sulfur atom. It has been proposed that the tryptophan modulates the electronic properties of the heme and the heme-bound ligands. DFT QM/MM calculations have been used to investigate the role of the tryptophan in the product state, i.e., NO bound to the heme iron (Fernandez et al. 2005). Several studies on the reaction mechanism of NOS have also been performed, with small "cluster" models with quantum mechanical (QM) electronic structure calculations and with ▶ QM/MM methods on larger molecular models incorporating the whole enzyme environment (de Visser 2009). These studies focus on the initial reaction step, the oxidation of L-arginine to *N*-hydroxo-arginine. It was suggested that due to the unique and highly polar L-arginine substrate, the catalytic cycle for the first step in NOS diverges from cytochrome P450 enzymes after formation of Compound I, and Compound I is not the active oxidant, in contrast to cytochrome P450 enzymes. Subsequent QM/MM calculations considered three different possible reaction mechanisms for the first step in NOS. Only the pathway that featured a singly protonated ferric-peroxy complex, combined with formation of a cation-radical species shared by the tetrahydrobiopterin cofactor and the deprotonated arginine substrate, was deemed energetically reasonable and could account for experimental data on the reaction (Cho et al. 2009). The detailed catalytic mechanism of this enzyme is not yet entirely resolved, but it is clear that computational molecular simulation and modeling provides a crucial means of proposing and testing proposed reaction pathways, because unambiguous experimental observation of the species involved is nigh impossible. A study into the nature of the tetrahydrobiopterin cofactor bound to NOS is an example of how experimental (e.g., EPR) and computational studies can be combined to help elucidate further details (Stoll et al. 2010).

Cross-References

▶ Cytochrome P450 – Computational Studies
▶ QM/MM Methods

References

Cho K-B, Carvajal MA, et al. First half-reaction mechanism of nitric oxide synthase: the role of proton and oxygen coupled electron transfer in the reaction by quantum mechanics/molecular mechanics. J Phys Chem B. 2009;113(1):336–46.
de Visser SP. Density functional theory (DFT) and combined quantum mechanical/molecular mechanics (QM/MM) studies on the oxygen activation step in nitric oxide synthase enzymes. Biochem Soc Trans. 2009;37:373–7.
Fernandez ML, Marti MA, et al. Proximal effects in the modulation of nitric oxide synthase reactivity: a QM-MM study. J Biol Inorg Chem. 2005;10(6):595–604.
Stoll S, Nejatyjahromy Y, et al. Nitric oxide synthase stabilizes the tetrahydrobiopterin cofactor radical by controlling its protonation state. J Am Chem Soc. 2010;132(33):11812–23.

NMR

▶ Assignment of ^{19}F Resonances in Protein Solution State NMR Studies
▶ Fragment Screen
▶ In-Cell NMR
▶ Multidimensional NMR Spectroscopy
▶ NMR in Drug Discovery – Introduction
▶ NMR Studies of Macromolecular Interactions – Introduction
▶ Nuclear Overhauser Effect
▶ Protein NMR Resonance Assignment
▶ Pulsed Field Gradient NMR
▶ SAR by NMR
▶ Stereo-Array Isotope Labeling (SAIL) Method
▶ Structure Determination by NMR: Overview
▶ Triple Resonance NMR

NMR in Drug Discovery – Introduction

Steven Halouska and Robert Powers
Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE, USA

Synonyms

Ligand-binding interactions; NMR; Target and hit validation techniques

Definition

Processes by which drugs are discovered and designed with analytical techniques such as NMR.

Introduction

Drug discovery is an ongoing and challenging process that is fraught with failure, but the successes have had profound impacts on human health. Fundamentally, drug discovery is a multidisciplinary endeavor requiring, among others, bioinformatics and computational chemistry, cell biology, medicinal chemistry, enzymology, high-throughput screening (HTS), molecular biology, protein chemistry, genomics and other "omics" technologies, and structural biology. Drug discovery is also an iterative procedure comprising four stages: target identification and validation, lead discovery, lead optimization, and clinical trials (Betz et al. 2006). In general, the drug discovery process starts by identifying a protein target associated with drug resistance, the progression or pathology of a disease, or the virulence of an organism. Usually, the protein of interest is chosen because it is perturbed when the system is in a disease state, or the target is a critical or essential protein, for instance, required for microbial or tumor survivability. Part of the target identification process also requires understanding the protein's biological or cellular function and potentially obtaining its three-dimensional (3D) structure. The process of finding an initial set of antagonists or agonists is typically accomplished using HTS and a library composed of hundreds of thousands to millions of compounds. Commonly, three to five chemical classes from the HTS results are selected as chemical leads for further optimization. The selection of chemical leads is based on activity, chemical novelty and patentability, drug-like characteristics, synthetic accessibility, and diversity. Different analogs of the chemical leads are synthesized to increase affinity, selectivity, and potency, while reducing toxicity issues. Effectively, the goal is to evolve the chemical leads into drug candidates. Overall, the drug discovery process is extremely time consuming and expensive, requiring approximately 12 years and costing >$800 million dollars. This is due, in part, to the complexity of the biological system, limited knowledge about the protein targets, and the inherent difficulty of converting an inhibitor to a drug, where only 1 in 5,000 chemical leads becomes a drug (Light and Warburton 2011). Therefore, technological advancements and new methodologies that improve the success rate are desperately needed.

Nuclear magnetic resonance (NMR) is a versatile analytical tool that is used in all phases of the drug discovery process prior to initiating clinical trials (Betz et al. 2006). NMR is routinely used to characterize the structures of both small molecules and large biomolecules. More importantly, NMR is often used to address an initial and important question in the drug discovery process: does the ligand selectively and specifically bind to the protein target in a biologically relevant manner? This is an essential step in evaluating the results obtained from HTS, in silico screens, or fragment-based screens, and for evolving chemical leads into drug candidates. NMR is uniquely suited for detecting protein-ligand interactions, for identifying the ligand-binding site, for calculating dissociation constants (K_D), for determining a 3D structure of the protein-ligand complex, and for monitoring in vivo activity. Furthermore, this information can be obtained using HTS-NMR, where hundreds to thousands of potential drugs can be screened. In addition, cryoprobe technology, high-field magnets (>800 MHz), and enhancement in NMR pulse sequences and HTS technology have substantially increased throughput, signal-to-noise, and the detection of low levels of proteins or weak binding ligands. Thus, HTS-NMR has been applied to metabolomics, and extended to larger chemical libraries and larger protein targets (>100 kDa), while simultaneously minimizing resources (experimental time, material) (Pellecchia et al. 2008).

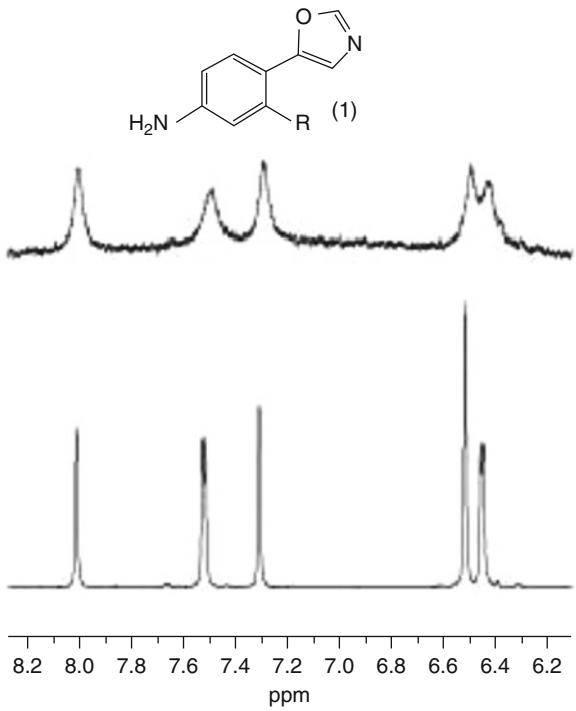
An inherent value of HTS-NMR, with respect to drug discovery, is its nearly universal application (Widmer and Jahnke 2004). HTS-NMR does not require any target-specific setup because the screen is independent of the protein's function. Thus, proteins recently identified by genomics can be screened immediately, as long as an adequate amount of labeled or unlabeled material is available; and the protein falls within the allowable molecular weight (MW) range for the specific NMR experiment. In general, the protein is overexpressed and isotopically labeled with ^{15}N and/or ^{13}C for structure determination or HTS affinity screens. *Escherichia coli* and *Bacillus subtillis* are the most commonly used expression systems because their robustness permit high cell density and expression rates (Heller and

Kessler 2001). The 3D protein structure is critical to the drug discovery process because it enables the evaluation of the biological relevance of the potential drug candidates. This is accomplished by determining if the ligand-binding site corresponds to the protein's active site or functional epitope. Similarly, ligand selectivity and specificity is determined based on the ligand binding to other binding sites or proteins. Also, the structure is routinely used to annotate functionally uncharacterized proteins based on structural homology to proteins with an assigned function. The NMR assignments and 3D protein structures are typically determined using a standard set of 3D triple resonance and NOESY experiments (Kanelis et al. 2001).

HTS-NMR is used to identify chemical leads, and validate HTS and in silico screening results. SAR by NMR (structure-activity relationship by NMR) was the first illustration of HTS-NMR. Since then, there have been numerous NMR experiments designed for the sole purpose of being used in an NMR ligand affinity screen (Pellecchia et al. 2002). Some popular one-dimensional NMR experiments (or ligand-focused screens), include saturation transfer difference (STD), WaterLOGSY (water ligand observation by gradient spectroscopy), NOE pumping, and diffusion edited NMR (Widmer and Jahnke 2004). The primary goal of these experiments is to use an observable change in the NMR spectrum of the chemical lead (peak intensity, peak width, chemical shift, etc.) to identify a productive binding interaction. Advantages of these techniques include rapid data collection (<5 min.), minimal sample requirements (<10 μM), and no need for isotopically labeled protein. But in general, these techniques do not provide any information on the ligand-binding site. Alternatively, protein-focused screens, SAR by NMR, along with Multi-Step NMR, MS/NMR, NMR-SOLVE, and SHAPES, among others, are used to identify ligand-binding sites in addition to identify binders. Importantly, these methods are typically used as part of a fragment-based screen that requires determining the binding proximity of two or more ligands. The goal is to chemically link multiple fragments to achieve a corresponding enhancement in binding affinity. But, these screens do require isotopically labeled proteins and significantly longer data acquisition times. HTS-NMR can also be used to evaluate in vivo drug toxicity and efficacy by using NMR metabolomics protocols. NMR is used to compare metabolic profiles before and after drug treatment to determine the biological impact of the drug candidate.

Ligand-Focused HTS-NMR

The relatively low sensitivity of NMR is an important issue in its application to HTS. Correspondingly, HTS-NMR requires longer data acquisition times and sample requirements compared to standard high-throughput screens. Importantly, screening hundreds of thousands to millions of compounds, while routine for HTS, is completely impractical for NMR. Thus, designing NMR experiments to minimize both instrument time and sample requirements is critical to improve throughput; and the primary goal of ligand-focused 1D NMR screens. Simply, a ligand-focused screen monitors changes in the 1D ^{1}H NMR spectrum of the ligand resulting from the addition of the protein target. The ligand is typically in large excess (>10-fold) of the protein. In general, the 1D ^{1}H-NMR methods exploit large differences in physical properties between small molecules and large biomolecules; and correspondingly between the free and bound states of the ligand. For example, a schematic representation in Fig. 1 illustrates the effects of differing T_2 relaxation times on ligand binding (Fejzo et al. 1999). Specifically, small molecules undergo rapid Brownian motion and have correspondingly slow T_2 relaxation times. Conversely, proteins undergo slow Brownian motion and have fast T_2 relaxation times. Therefore, the NMR spectrum of a small molecule consists of narrower peaks compared to a protein NMR spectrum. As a small molecule binds a protein, it will inherit the physical properties of the protein. As a result, the NMR peaks of the small molecule will broaden proportionally to the difference in line widths and fraction of bound ligand. Ligand-focused 1D NMR screens have some distinct advantages. The NMR experiments are relatively fast (<5 min). Only small quantities, ~1–10 μM of unlabeled protein are required per NMR experiment. Additionally, the ligands can be screened as mixtures of ~4 to >20 different compounds with a concentration of ~100 μM per ligand. Assembling large mixtures can be challenging since it may be difficult to find compatible sets of compounds that maintain solubility and stability while avoiding cross-reactivity.



NMR in Drug Discovery – Introduction, Fig. 1 1D ^{1}H spectra of 1 mM free ligand (*bottom*) in the presence of 100 μM IMP dehydrogenase (*top*). The significant line broadening indicates a ligand-protein interaction. (Reprinted with permission from (Fejzo et al. 1999). ©1999 by Elsevier)

The appearance of the 1D ^{1}H NMR spectrum also reflects on how rapidly the ligand exchanges between the bound and free state. Additionally, the appearance of the 1D ^{1}H NMR spectrum is influenced by the ligand's affinity to the protein target (Pellecchia et al. 2002):

$$K_D = \frac{K_{off}}{K_{on}} = \frac{[L]_F[P]_F}{[PL]} \tag{1}$$

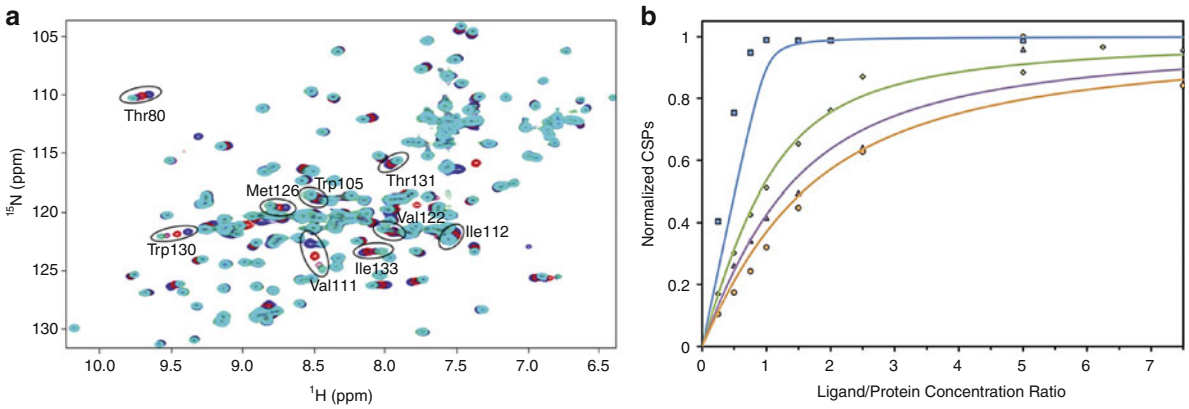
where K_D is the dissociation constant, K_{off} is the rate constant for dissociation, K_{on} is the rate constant for formation, $[L]_F$ is the concentration of free ligand, $[P]_F$ is the concentration of the free target protein, $[PL]$ is the concentration of the protein-ligand complex. Tight binders usually have slow exchange rates and a K_D of less than 10 nM. In this case, a separate signal will be observed for both the bound and free ligands, where the bound signal is usually broadened beyond detections. Correspondingly, a binding event is indicated by a decrease in the intensity of the ligand NMR spectrum. If the ligand binds weakly ($K_D > 10$ μM), the exchange between the bound and free state is fast and the observed spectrum consists of a single signal, whose chemical shift and line width is the weighted average between the bound and free state. For intermediate exchange, the ligand peak broadening is typically enhanced. Thus, in the case of fast exchange, binding affinities can be estimated from HTS-NMR using 1D ^{1}H NMR spectra (Shortridge et al. 2008). The K_D can be calculated by measuring the peak height and line width using the following equations:

$$K_D = \left[\left(\frac{c[P]_T}{B_{\text{single}}} - c[P]_T\right) - [L]_T\right]$$
$$\text{where } c = \frac{\upsilon_B}{\upsilon_F} - 1 \text{ and } B_{\text{single}} = 1 - \frac{I_B}{I_F} \tag{2}$$

where $[P]_T$ and $[L]_T$ are the total protein and ligand concentrations, respectively, v_B and v_F are the line width for the bound and free state, respectively, and I_B and I_F are the peak height for the bound and free state, respectively. In addition to fast exchange, the approach also assumes that there is only a single binding site. These assumptions are reasonable in the early stages of drug discovery because initial chemical leads will generally have weak binding affinities. Obtaining binding affinities from HTS-NMR have also been described for STD experiments, for the displacement of known low-affinity inhibitors, or for chemical shift changes using ^{19}F NMR with ^{19}F-containing compounds, among other NMR experiments.

Target-Focused HTS-NMR

One important advantage of HTS-NMR is the ability to detect weak binding ligands, which is a critical factor in fragment-based screens. Most target-focused approaches to HTS-NMR rely on chemical shift perturbations in the protein NMR spectrum, where the protein is either ^{15}N or ^{13}C labeled. The ligand is unlabeled and undetected by the target-based NMR experiments. Therefore, the ligand is still typically in excess (≥5-fold) to maximize the observation of weak binding ligands ($K_D > 1$ mM). Two-dimensional (2D) ^{1}H-^{15}N-heteronuclear single quantum coherence (^{1}H-^{15}N-HSQC) experiments are commonly used for HTS-NMR. Other common HTS-NMR experiments include the 2D ^{1}H-^{15}N-TROSY, 2D ^{1}H-^{13}C HSQC, and recent derivatives that enable very



NMR in Drug Discovery – Introduction, Fig. 2 (**a**) Overlay of the 2D ^{1}H-^{15}N-HSQC spectra of *B. subtilis* protein YndB with an increasing concentration of chalcone ranging from 0 μM (*blue*) to 160 μM (*cyan*). (**b**) NMR titration data for *trans*-chalcone (*blue*), flavanone (*green*), flavone (*purple*), and flavonol (*orange*). The normalized chemical shift perturbations (CSP) for the nine most perturbed residues are plotted versus the protein-ligand concentration ratios. The theoretical curve displayed for *trans*-chalcone corresponds to a K_D of 1 μM and represents the upper limit for the K_D. The measure K_D values are ≤1 μM (*trans*-chalcone), 32 ± 3 μM (flavanone), 62 ± 9 μM (flavone), and 86 ± 16 μM (flavonol). (Reprinted with permission from (Stark et al. 2010). © 2010 by John Wiley and Sons)

rapid data collection (e.g., SOFAST). A peak in the 2D ^{1}H-^{15}N-HSQC spectrum is correlated to each amino acid residue (except proline) present in the protein sequence. Each peak is also associated with both a ^{1}H and ^{15}N chemical shift resulting from the bonded NH (amide) group. The relative location (chemical shifts) of each peak in the 2D ^{1}H-^{15}N-HSQC spectrum is strongly dependent on the local environment for each amino acid residue in the protein structure. Thus, the addition of a ligand known to bind the protein results in local environmental changes for amino acids in direct contact with the bound ligand, and for residues that undergo an induced structural change. Correspondingly, chemical shift changes are observed for a select number of peaks that can be mapped onto the protein structure to identify the ligand-binding site. A unique spatial clustering of residues that incur a chemical shift perturbation (CSP) indicates a specific binding interaction and the location of the ligand-binding site. Conversely, a random distribution of residues with CSPs on the proteins surface indicates a nonspecific interaction. Alternatively, an observation that a majority of residues incur a CSP or experience a decrease in peak intensity probably infers a ligand-induced aggregation, precipitation, or denaturation, which typically eliminates the compound's consideration as a chemical lead. Importantly, the observation that two or more ligands (fragments) share proximal, but not identical binding sites, based on CSP mapping, presents an opportunity to chemically link the fragments. The linked fragments are expected to have an enhanced affinity and represent a starting point for further optimization.

The addition of an increasing concentration of unlabeled ligand will cause a progressive change in the protein's chemical shifts if the ligand-binding affinity is weak and in fast exchange (Roberts 2000). Thus, a binding affinity can be determined by following CSPs as a function of ligand concentration:

$$CSP_{obs} = \frac{(K_D + [L] + [P]) - \sqrt{(K_D + [L] + [P])^2 - (4[L][P])}}{2[P]} \quad (3)$$

where CSP_{obs} is the 2D ^{1}H-^{15}N HSQC chemical shift perturbations, $[P]$ is the protein concentrations, $[L]$ is the ligand concentration, and K_D is the dissociation constant. Conversely, a tightly bound ligand in slow exchange will result in two sets of NMR peaks in the HSQC spectrum for the bound and free state. The ratio of intensities for the bound and free states can be used to determine the fraction bound and measure a K_D from a standard binding isotherm. Figure 2 illustrates the application of a 2D ^{1}H-^{15}N HSQC titration experiment to measure chemical shift perturbations and calculate a K_D (Stark et al. 2010).

HTS-NMR and Lead Optimization

It has been estimated that the number of possible chemical compounds is on the order of 10^{60} (Bohacek et al. 1996). Clearly, obtaining and experimentally screening such an immense compound library is impossible. Instead, a screening library for HTS-NMR usually consists of only a few hundred to several thousand compounds. Commonly, the HTS-NMR screening library is a set of chemical leads from HTS. Obviously, the quality of the chemical leads that emerge from HTS is dependent on the quality of the original HTS library (Lipinski 2004). A major source of populating a screening library is from prior HTS lead optimization. Unfortunately, this often leads to higher MW and more lipophilic compounds due to exploiting hydrophobic interactions that increase potency. The result is a compound library with reduced pharmacokinetics or pharmacodynamics properties. In essence, HTS-NMR provides an invaluable function by further validating the HTS chemical leads and experimentally eliminating compounds with poor drug-like properties. HTS-NMR can experimentally verify acceptable physical behavior, such as good solubility, stability, purity, and a lack of aggregation or micelle-like behavior. HTS-NMR can also confirm that the chemical lead specifically binds the protein target in a biologically relevant manner. Additionally, selection of compounds can also abide by the Lipinski's rule of 5, which is based on predicted molecular properties important for drug-like behavior. Of course, strictly following Lipinski's rule of 5 may eliminate potentially interesting and novel leads. Instead, a combination of HTS-NMR and Lipinski's rule of 5 provides the most flexibility. Additionally, ligand efficiency (LE) provides a better approach to prioritize chemical leads instead of traditional measures of relative activity (IC_{50}, or K_i). LE is defined as:

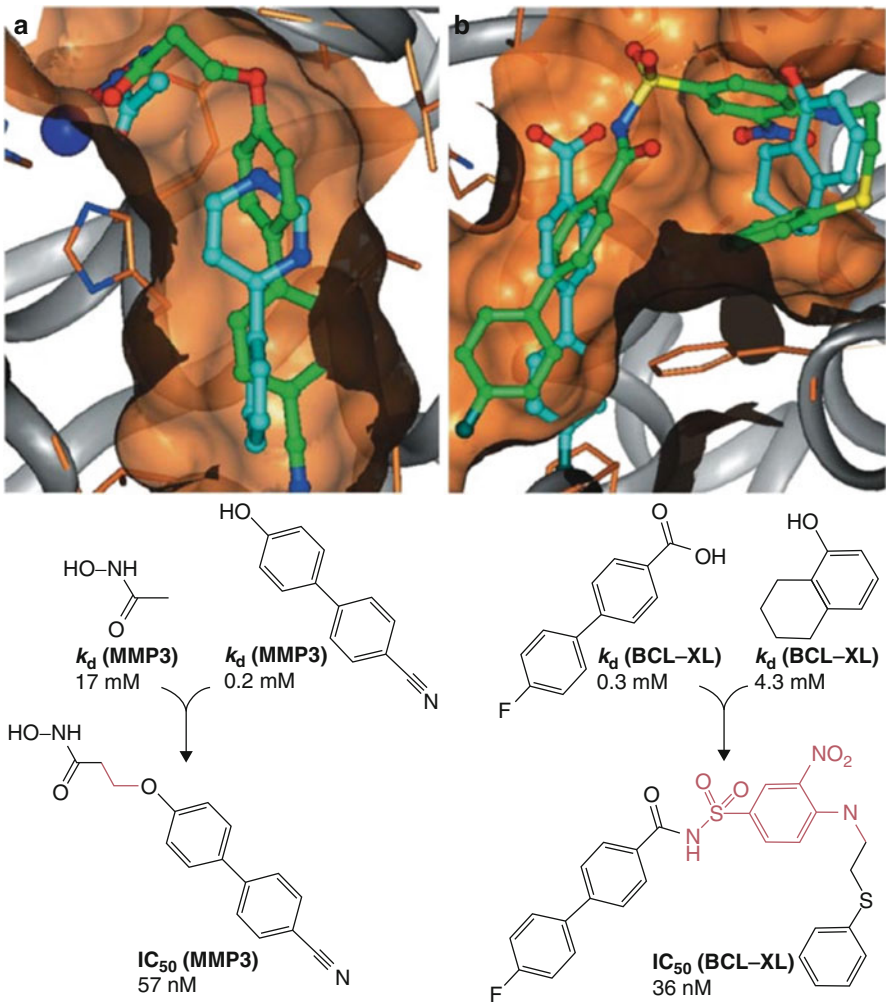
$$LE = \frac{\Delta G}{N} = \frac{-RT \ln K_i}{N} \tag{4}$$

where ΔG is the Gibbs free energy, K_i is the dissociation constant for inhibitor binding, and N is the number of non-hydrogen atoms.

Based on LE, lower MW (<300 Da) chemical leads provide a better opportunity to successfully evolve a compound into a drug (Zartler and Shapiro 2005). Similarly, a library of low MW compounds provides a more efficient coverage of structural space because the number of low MW compounds is significantly reduced. The efficiency of low MW compounds for drug discovery is the basis of fragment-based screening. Correspondingly, HTS-NMR using a fragment-based library has become an integral part of the drug discovery process. A fragment-based screen is complementary to standard HTS because the two chemical libraries typically cover different regions of structural space. As a result, chemical leads from a fragment-based screen tend to be novel compounds and distinct from HTS leads. Also, chemical leads from a fragment-based screen tend to be more drug-like, where the lower MW provides more opportunities for optimization and maintaining high ligand efficiency. Again, the primary goal behind a fragment-based screen is to identify two or more fragments that bind in proximal sites on the target protein. Chemically linking these fragments in a manner that maintains the original binding orientations will result in a new lead molecule with enhanced affinity. The process is illustrated in Fig. 3. HTS-NMR is well-suited to screening a fragment-based library because the low MW drug-like fragments are expected to have weak binding affinities (K_D > μM – mM) that are readily detectable by NMR. Patentability may be a potential challenge with fragment-based libraries. Multiple pharmaceutical and biotechnology companies will use very similar screening libraries since chemical fragments of known drugs are a well-defined set of compounds. Thus, it will not be surprising to obtain similar chemical leads when fragment-based libraries are screened against an identical protein target.

NMR Functional Annotation

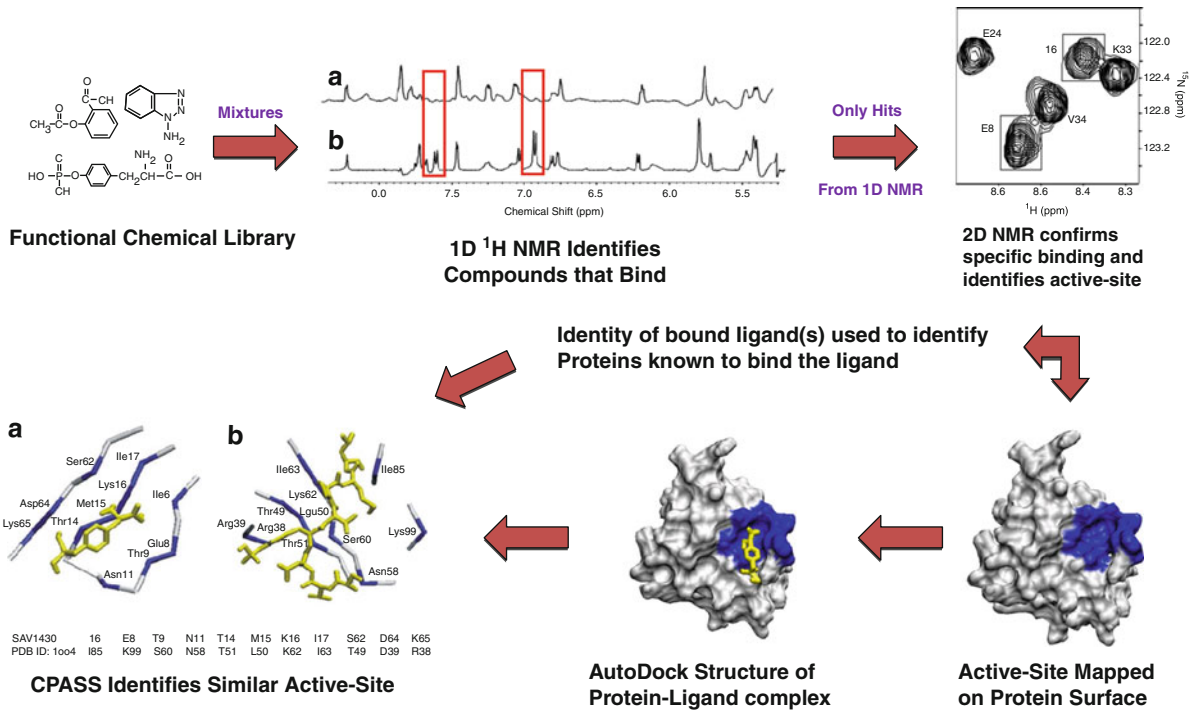
Functional annotation of unknown protein targets plays an important role in the drug discovery process because it allows for identification of novel targets that may be involved in human disease. Functional annotation is primarily obtained by sequence or structure homology. The majority of the ~19 million proteins identified to date (UniPortKB; http://www.uniprot.org) are functionally annotated through sequence or structure homology. The transfer of functional information is vastly more common than the direct acquisition of experimental data. But at least 40% of prokaryotic and eukaryotic genomes are explicitly annotated as "hypothetical" or "uncharacterized" proteins because of a lack of



NMR in Drug Discovery – Introduction,
Fig. 3 Applications of SAR by NMR method for fragment design showing ligand binding for (**a**) matrix metalloproteinase 3 (**b**) Bcl-xL. In each case at the *top*, the identified fragment leads are shown with *cyan* carbons, whereas the linked compounds are denoted with *green* carbon atoms. (Reprinted with permission from (Hajduk and Greer 2007). © 2007 by Nature Publishing Group)

homology to functionally annotated proteins. HTS-NMR can be used to assist in the functional classification of unannotated proteins when sequence and structure homology fails.

A fundamental component to understanding a protein's function is derived from its interaction partners. The identity of ligands and the location of binding sites can be leveraged to infer a function in a manner similar to sequence and structure homology. This is based on the observation that amino acids in an active site or functional epitope are evolutionarily more stable than the rest of the protein. Effectively, these residues are required to maintain function, where proteins that share similar ligand-binding sites are predicted to be functional homologs. FAST-NMR (**F**unctional **A**nnotation **S**creening **T**echnology using NMR) uses HTS-NMR to identify ligands and binding sites for functionally uncharacterized proteins. FAST-NMR then exploits a structure and sequence similarity to a ligand-binding site from an annotated protein to infer a function for an uncharacterized protein (Powers et al. 2008). A schematic representation of the FAST-NMR protocol is illustrated in Fig. 4. Unlike traditional HTS-NMR, the screening library contains only biologically active molecules that are active against a defined protein target. In essence, the library is used as chemical probes to identify structural homologs of the natural ligand and the location of ligand-binding sites. FAST-NMR uses a tiered approach to HTS-NMR that includes a 1D ^{1}H NMR ligand-focused screen followed by a 2D ^{1}H-^{15}N-HSQC target-focused CSP screen. The ligand-focused screen is used to quickly identify binders, where only the hits are further screened in the target-focused CSP screen. The 2D ^{1}H-^{15}N-HSQC CSP screen is used to verify binders, identify the



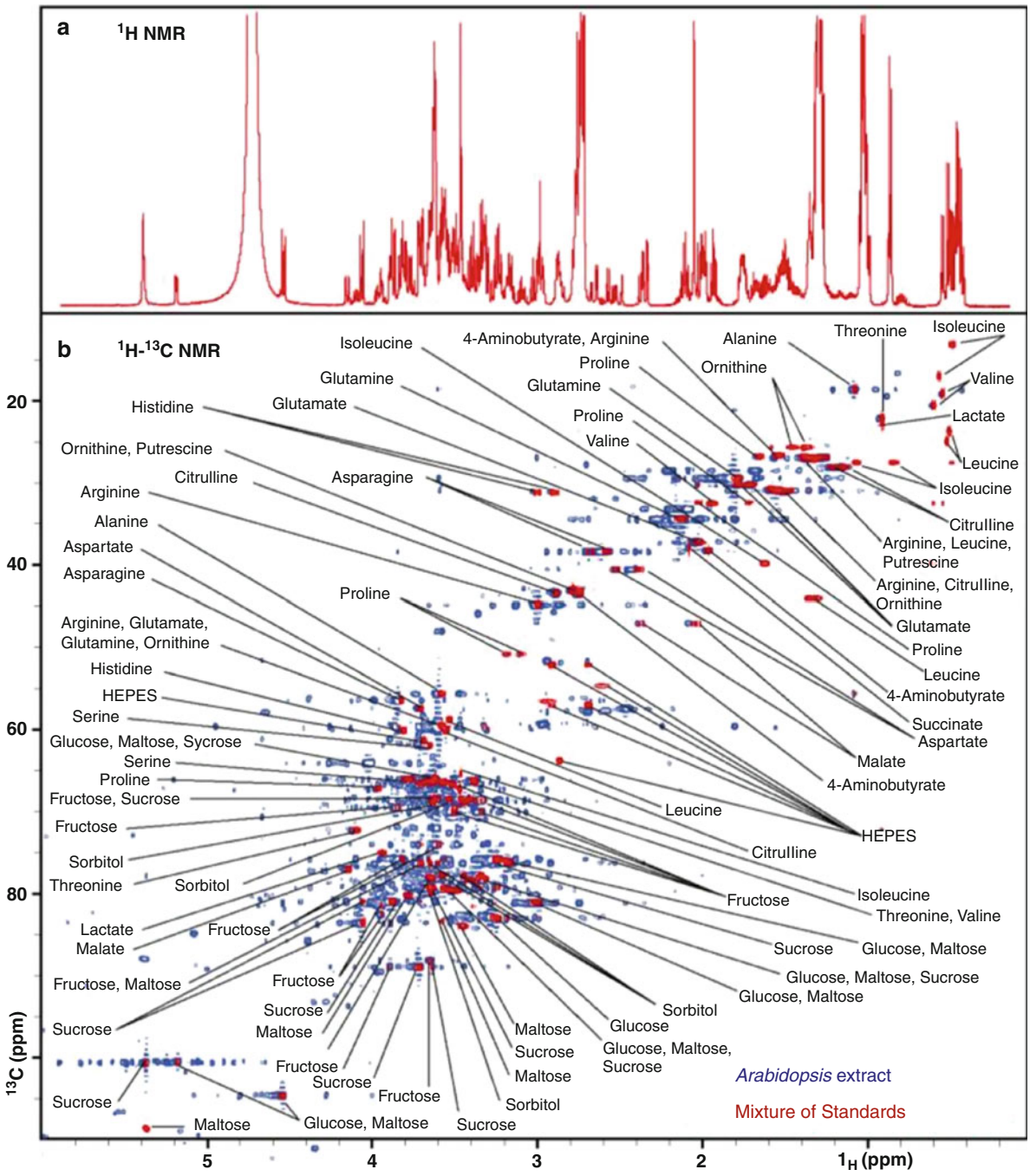
NMR in Drug Discovery – Introduction, Fig. 4 Schematic diagram for FAST-NMR. (Reprinted with permission from (Powers et al. 2008). © 2008 by Elsevier)

ligand-binding site, and generate a protein-ligand co-structure. This structure is then used as an input for CPASS (**C**omparison of **P**rotein **A**ctive-**S**ite **S**tructure) to identify homologous ligand-binding sites from functionally annotated proteins. The CPASS database contains all the unique ligand-binding sites present in the RCSB PBD.

NMR Metabolomics and In Vivo Drug Activity

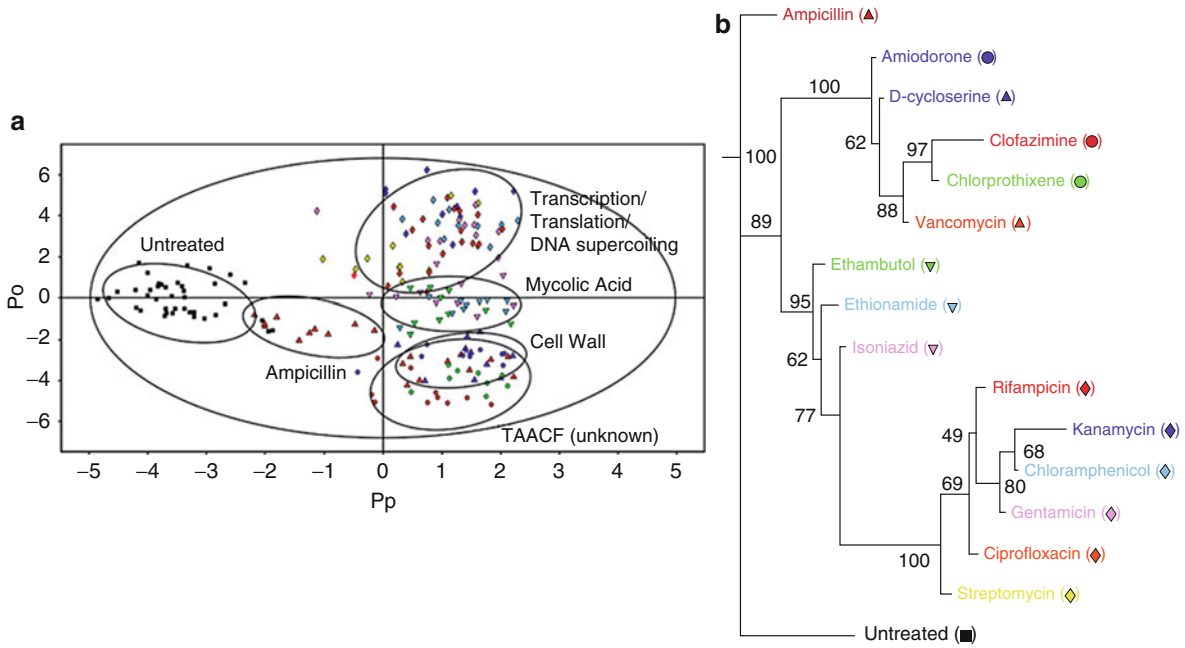
One means of analyzing the state of a biological system is achieved by monitoring the metabolome – all the metabolites present in a cell, tissue, organ, or organism. Correspondingly, metabolomics is the study of the changes in the concentration and identity of these metabolites that results from environmental or genetic stress from a disease state or drug treatment. In essence, the metabolome provides a chemical fingerprint or signature that uniquely defines the state of the system. Metabolomics has an intrinsic advantage over genomics and proteomics analysis since observed changes in the metabolome are directly coupled with changes in protein activity and cell function. Metabolites are the end product of all cellular processes, and are a direct result of enzymatic and protein activity. NMR metabolomics can assist drug discovery by characterizing drug efficacy, selectivity, and toxicity (Powers 2009). Importantly, NMR metabolomics may identify and eliminate troublesome compounds prior to initiating a clinical trial. Thus, HTS-NMR can also be used to follow the impact of chemical leads on metabolites from cellular extracts (bacteria, human cell lines, tumors, etc.) or biofluids (urine, serum, saliva, etc.). Common NMR experiments used for metabolomics includes: 1D ^{1}H NMR, 2D ^{1}H-^{1}H TOCSY, and 2D ^{1}H-^{13}C HSQC.

Metabolites are typically detected using 1D ^{1}H-NMR, where peak heights relative to an internal standard are used to determine metabolite concentrations. Because of the low sensitivity of NMR, only the most abundant ($\sim$ >1 μM) metabolites are observed. Also, a metabolomic sample is a complex heterogeneous mixture containing numerous metabolites. The limited spectral resolution and severe peak overlap in a 1D ^{1}H-NMR spectrum makes unambiguous metabolite



NMR in Drug Discovery – Introduction, Fig. 5 (**a**) One-dimensional ^{1}H-NMR spectrum of an equimolar mixture of the 26 small-molecule standards. (**b**) Two-dimensional ^{1}H-^{13}C-HSQC NMR spectra of the same synthetic mixture (*red*) overlaid onto a spectrum of aqueous whole-plant extract from *Arabidopsis thaliana* (*blue*). (Reprinted with permission from (Lewis et al. 2007). © 2007 by American Chemical Society)

assignment particularly challenging. Furthermore, since the metabolome has not been completely characterized, erroneous assignments may result. Alternatively, 2D ^{1}H-^{13}C HSQC spectra are commonly used to identify metabolites. The higher spectral resolution, the larger ^{13}C chemical shift dispersion, and the observation of



NMR in Drug Discovery – Introduction, Fig. 6 (**a**) 2D OPLS-DA scores plot demonstrating the clustering pattern for 12 antibiotics with known biological targets and three compounds of unknown in vivo activity: untreated *Mycobacterium smegmatis* cells, chloramphenicol, ciprofloxacin, gentamicin, kanamycin, rifampicin, streptomycin, ethambutol, ethionamide, isoniazid, ampicillin, D-cycloserine, vancomycin, amiodorone, chlorprothixene, and clofazimine treated *M. smegmatis* cells. The symbols correspond with the coloring scheme and labeled symbols indicated on the tree diagram in (**b**). The ellipses correspond to the 95% confidence limits from a normal distribution for each cluster. The untreated *M. smegmatis* cells (*black square*) was designated the control class, and the remainder of the cells were designated as treated. The OPLS-DA used one predictive component and six orthogonal components to yield a R^2X of 0.715, R^2Y of 0.803, and Q^2 of 0.671. (**b**) Metabolomics tree diagram determined from the OPLS-DA scores plot. The coloring scheme and associated symbol for each compound in the tree diagram correlates with colored symbols in the OPLS-DA scores plot. The bootstrap numbers for each node are indicated on the tree diagram. (Reprinted with permission from (Halouska et al. 2012). © 2011 by American Chemical Society)

chemical shifts for each C-H pair significantly improve the accuracy of metabolite assignment. The metabolomic sample does need to be supplemented with a ^{13}C-labeled metabolite to avoid long acquisition times because of the low natural abundance of ^{13}C (1.1%). Correspondingly, the only metabolites observed in the 2D ^{1}H-^{13}C HSQC spectrum are derived from the ^{13}C-labeled metabolite. Also, because of variations in J-coupling constants, and T_1 and T_2 relaxation times, direct quantitation of a standard HSQC spectrum to calculate metabolite concentrations is unreliable. Nevertheless, modification to the HSQC NMR pulse sequence has produced quantitative HSQC experiments (e.g., $HSQC_0$, QQ-HSQC, etc.). Metabolites can also be quantified by using the fast quantification by NMR method (FMQ), where a series of standard metabolites over a range of concentrations are screened (Fig. 5). A biological extract can then be compared against the standards to determine the metabolite concentrations (Lewis et al. 2007).

Several databases have been developed that enable metabolite identification from experimental chemical shift data. The databases contain NMR and mass spectroscopy data for thousands of metabolites from a broad range of organisms. Therefore, metabolite identification needs to be verified with KEGG and MetaCyc metabolic pathway databases to verify the metabolite is actually present in a specific organism (Powers 2009). The NMR metabolomic databases include: Human Metabalome Database (HMDB), Madison Metabolomics Database (MMCD), and Platform for Riken Metabolomics (PRIMe).

A common application of NMR metabolomics data is to monitor global differences between metabolic samples instead of following changes in specific

metabolites. This approach involves the use of multivariate statistical analysis and pattern recognition. Multivariate statistical techniques such as principal component analysis (PCA) and orthogonal partial least-squares discriminant analysis (OPLS-DA) are routinely employed to capture global perturbations in the metabolome (Stoyanova and Brown 2002; Bylesjo et al. 2006). PCA is an unsupervised method that reduces multivariate dataset to a single point and projects the major variations in the dataset into a few axes called principal components. In this way, spectral variations are captured in a model that can be easily visualized. As a result, similar metabolomic data will cluster closely together and separately from other distinct metabolomes. Conversely, OPLS-DA is a supervised method used to determine the variations within the data set that is correlated to the classification label (treatment, control, disease state). If the variations within the data set (noise) are not correlated with the classification label, then the variations are filtered out, resulting in a single latent vector. This is analogous to PCA, where PCA captures all the intrinsic variations within the data set. Applying PCA and OPLS-DA to NMR metabolomics data sets is extremely valuable for classifying drug activity and toxicity in a living system based on relative clustering patterns (Fig. 6). This also allows for determining the in vivo mechanism of action for a new chemical lead or drug (Halouska et al. 2012).

Summary

Nuclear magnetic resonance (NMR) is a versatile analytical tool with a wide range of applications that may improve the success rate of drug discovery. NMR uses ligand-focused or target-focused screening techniques to discover chemical leads with a large range of affinities. Correspondingly, NMR is ideally suited for fragment-based screening, which complements standard high-throughput screens and expands the coverage of compound structural space. NMR can also be employed to identify the location of ligand-binding sites and determine a protein-ligand co-structure. This is essential information for linking fragments to enhance binding affinity and to evolve chemical leads into drug candidates. NMR ligand affinity screens can also be applied to assist in the functional annotation of uncharacterized proteins to identify new therapeutic targets. Finally, NMR metabolomics strategies can identify the in vivo mechanism of action, and determine the in vivo efficacy and toxicity for chemical leads and help alleviate the many failures encountered in clinical trials.

Cross-References

- ▶ BMRB
- ▶ Flow NMR
- ▶ Fragment Screen
- ▶ Mass Spectrometry: Application to Protein-Ligand Interactions
- ▶ Metabolomics
- ▶ Multidimensional NMR Spectroscopy
- ▶ NMR
- ▶ NMR-based Structural Proteomics
- ▶ Protein NMR – Introduction
- ▶ Protein Secondary Structure Prediction in 2012
- ▶ SAR by NMR
- ▶ SOFAST HMQC
- ▶ Total Correlation Spectroscopy (TOCSY) in NMR Protein Structure Determination
- ▶ Triple Resonance NMR
- ▶ TROSY

References

Betz M, Saxena K, et al. Biomolecular NMR: a chaperone to drug discovery. Curr Opin Chem Biol. 2006;10(3):219–25.

Bohacek RS, McMartin C, et al. The art and practice of structure-based drug design: a molecular modeling perspective. Med Res Rev. 1996;16(1):3–50.

Bylesjo M, Rantalainen M, et al. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. J Chemom. 2006;20(8–10):341–51.

Fejzo J, Lepre CA, et al. The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. Chem Biol. 1999;6(10):755–69.

Hadjduk J, Greer J. A decade of fragment-based drug design: stratigeic advances and leassons learned. Nat Rev Drug Discov. 2007;6:211–19.

Halouska S, Fenton RJ, et al. Predicting the in vivo mechanism of action for drug leads using NMR metabolomics. ACS Chem Biol. 2012;7(1):166–71.

Heller M, Kessler H. NMR spectroscopy in drug design. Pure Appl Chem. 2001;73(9):1429–36.

Kanelis V, Forman-Kay JD, et al. Multidimensional NMR methods for protein structure determination. IUBMB Life. 2001;52(6):291–302.

Lewis IA, Schommer SC, et al. Method for determining molar concentrations of metabolites in complex solutions from

two-dimensional 1H-13C NMR spectra. Anal Chem. 2007;79(24):9385–90.
Light DW, Warburton R. Demythologizing the high costs of pharmaceutical research. Biosocieties. 2011;6(1):34–50.
Lipinski C. Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol. 2004;1(4):337–41.
Pellecchia M, Sem DS, et al. NMR in drug discovery. Nat Rev Drug Discov. 2002;1(3):211–9.
Pellecchia M, Bertini I, et al. Perspectives on NMR in drug discovery: a technique comes of age. Nat Rev Drug Discov. 2008;7(9):738–45.
Powers R. NMR metabolomics and drug discovery. Magn Reson Chem MRC. 2009;47(Suppl 1):S2–11.
Powers R, Mercier KA, et al. The application of FAST-NMR for the identification of novel drug discovery targets. Drug Discov Today. 2008;13(3–4):172–9.
Roberts GCK. Applications of NMR in drug discovery. Drug Discov Today. 2000;5(6):230–40.
Shortridge MD, Hage DS, et al. Estimating protein-ligand binding affinity using high-throughput screening by NMR. J Comb Chem. 2008;10(6):948–58.
Stark JL, Mercier KA, et al. Solution structure and function of YndB, an AHSA1 protein from *Bacillus subtilis*. Proteins. 2010;78(16):3328–40.
Stoyanova R, Brown TR. NMR spectral quantitation by principal component analysis. III. A generalized procedure for determination of lineshape variations. J Magn Reson. 2002; 154(2):163–75.
Widmer H, Jahnke W. Protein NMR in biomedical research. Cell Mol Life Sci CMLS. 2004;61(5):580–99.
Zartler ER, Shapiro MJ. Fragonomics: fragment-based drug discovery. Curr Opin Chem Biol. 2005;9(4):366–70.

NMR Ligand Affinity Screen

► Fragment Screen

NMR Methods for Kinetic Analysis

Igor L. Barsukov
School of Biological Sciences,
University of Liverpool, Liverpool, UK

Definition

NMR kinetic analysis aims at deriving reaction rates from the characteristics of the NMR spectra.

Basic Characteristics

The unique ability of NMR to resolve and monitor signals of individual nuclei in different chemical environment makes it a powerful method for kinetic analysis. NMR spectra are highly sensitive to the environment, and no modification of the molecules is normally required to detect the changes due to the interaction and kinetic effects. Its application, however, is limited by the low sensitivity and relatively long time of data acquisition. These restrictions are particularly severe for monitoring nonequilibrium processes where reaction course is followed after system perturbation or mixing, making these experiments generally suitable only for reaction times longer than seconds. Faster rates can be obtained by the NMR analysis of dynamic equilibrium where no net changes occur, although the reaction is proceeding at a detectable rate.

Nonequilibrium Analysis

The fastest way to monitor a reaction time course is through a sequential collection of 1D NMR spectra. A single scan acquisition takes a fraction of a second to complete and, even on high-field spectrometers, sufficient signal-to-noise ratio requires at least tens of scans at sub-millimolar concentrations. This restricts the delay between each point of the time course to seconds or longer. Additionally, there is a dead time before the experiment caused by sample insertion, temperature equilibration, and magnetic field adjustment. In a normal NMR setup, the dead time is measured in minutes, although can be dramatically reduced with the use of flow probes. Due to these considerations, even in an optimized setup at high concentrations, the measurable rate constants are less than 0.1 s^{-1}. For enzymatic reactions, the rate of substrate conversion can be easily adjusted by reducing enzyme concentration, making these reactions amenable to the nonequilibrium NMR analysis.

With a slow enough reaction rates, NMR provides a wealth of information for characterizing reaction pathways and intermediates. Signals of small molecules are usually well resolved in a 1D spectrum, and individual reaction products can be identified and followed selectively by monitoring corresponding resonances. If 1D resolution is not sufficient, a time series of 2D COSY or TOCSY spectra can be collected to improve resonance separation and help with product identification. Each of the 2D spectra normally takes tens of minutes to acquire, reducing the reaction rates available to the analysis. For ^{13}C or ^{15}N enriched molecules, corresponding HSQC spectra are used to enhance selectivity and resolution. Routinely, a sequential series